The Maximum Asymptotic Bias of Outlier Identifiers

Claudia Becker and Ursula Gather[1]

Abstract

In their paper, Davies and Gather (1993) formalized the task of outlier identification, considering also certain performance criteria for outlier identifiers. One of those criteria, the maximum asymptotic bias, is carried over here to multivariate outlier identifiers. We show how this term depends on the respective biases of estimators which are used to construct the identifier. It turns out that the use of high-breakdown robust estimators is not sufficient to achieve outlier identifiers with bounded maximum asymptotic bias.

Key words: Outlier identification; Robust statistics; Consistency.

1 Introduction

The performance of outlier identification rules cannot be judged by only one criterion. This is immediately clear by imagining how many different mechanisms may have created the 'outliers' and hence under how many situations our methods have to be compared w.r.t. their capability of labelling the right observations as outliers (cf. Barnett, Lewis, 1994, p. 121 ff). Important performance criteria describe e.g. the breakdown behaviour of such rules as a worst-case plot, where such descriptions include the finite-sample breakdown points of estimators used in these rules in the sense of Donoho and Huber (1983) as well as masking and swamping breakdown points of the identification procedures themselves as introduced by Davies and Gather (1993). These criteria concentrate on the question how large the amount of 'bad observations' in a sample has to be before the identification procedure under consideration breaks down in some sense. In this paper, we suggest a supplementary criterion, the maximum asymptotic bias, which has been defined for outlier identifiers by Davies and Gather in the univariate setting. The formal concept of outlier identification

[1]This work was supported by the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 475.

itself in the α outlier framework has already been extended to the multivariate situation in Gather and Becker (1997). Here, we give a definition of the maximum asymptotic bias of multivariate procedures, making use of a definition by Tyler (1994), who regards pairs of estimators for location and covariance. In contrast to breakdown criteria, the maximum asymptotic bias reflects the behaviour of an outlier identifier under a certain fixed proportion of 'bad observations' in the data. We derive necessary and sufficient conditions for identifiers with bounded maximum asymptotic bias. It turns out that using robust estimators with bounded bias in such identification procedures is only necessary, whereas a sufficient condition demands consistent estimators with $\sqrt{N}$ convergence rate.

This paper is organized as follows. In the following section, we recall the basic definitions of α outliers and outlier identifiers for the multivariate normal model. In Section 3, we present the definition of the maximum asymptotic bias, starting with Tyler's approach for a pair of estimators and adapting this notation to the necessities of outlier identifiers. Section 4 contains the results on necessary and sufficient conditions for identifiers with bounded maximum asymptotic bias.

2 Preliminaries

Aiming at the identification of outliers, we have to start with defining what we understand by this task. Davies and Gather (1993) introduced the approach to "define outliers in terms of their position relative to the model for the good observations", which leads to the concept of α outliers. Extending this to the situation of a multivariate normal distribution $N(\underline{\mu}, \Sigma)$, $\underline{\mu} \in I\!R^p$, $\Sigma \in I\!R^{p \times p}$ p. d., as the model distribution, Gather and Becker (1997) define an α *outlier* with respect to $N(\underline{\mu}, \Sigma)$ as an element of the α *outlier region*

$$\text{out}(\alpha, \underline{\mu}, \Sigma) := \{\underline{x} \in I\!R^p : (\underline{x} - \underline{\mu})^T \Sigma^{-1} (\underline{x} - \underline{\mu}) > \chi^2_{p;1-\alpha}\}.$$

For a sample of size N, the respective idea of an α_N outlier region $\text{out}(\alpha_N, \underline{\mu}, \Sigma)$ with $\alpha_N = 1 - (1-\alpha)^{1/N}$ can be given, such that

$$\mathrm{P}_{N(\underline{\mu},\Sigma)}(\underline{X} \in \text{out}(\alpha, \underline{\mu}, \Sigma)) = \alpha$$

and

$$\mathrm{P}_{N(\underline{\mu},\Sigma)}(\underline{X}_i \notin \mathrm{out}(\alpha_N, \underline{\mu}, \Sigma), i = 1, \ldots, N) = 1 - \alpha$$

for some $\alpha \in (0,1)$ in each case.

Knowing the outlier region, one can of course identify all α outliers. However, the parameters $\underline{\mu}$ and Σ of the model distribution are usually unknown, which means that we have to estimate this region. But, since we want to identify all α_N outliers in the sample, we already anticipate that this estimation must be based on a possibly corrupted sample, not coming i.i.d. from the model distribution. In any case, the estimation of the outlier region from a sample $\underset{\sim}{x}_N = (\underline{x}_1, \ldots, \underline{x}_N)$ yields an empirical or estimated outlier region

$$\underline{\mathrm{OR}}(\underset{\sim}{x}_N, \alpha_N) := \{\underline{x} \in I\!R^p : (\underline{x} - \underline{m})^T S^{-1} (\underline{x} - \underline{m}) \geq c\}.$$

The set $\underline{\mathrm{OR}}$ is also called an α_N *outlier identifier*, because every sample element lying in $\underline{\mathrm{OR}}$ can be understood as identified as an outlier in the sample at hand. Here, $\underline{m} = \underline{m}(\underset{\sim}{x}_N) \in I\!R^p$ and $S = S(\underset{\sim}{x}_N) \in I\!R^{p \times p}$, positive definite and symmetric, estimate $\underline{\mu}$ and Σ, respectively, and $c = c(p, N, \alpha_N) \in I\!R$ is a normalizing constant, calculated from some normalizing condition such as

$$\mathrm{P}_{N(\underline{\mu},\Sigma)}(\underline{X}_i \notin \underline{\mathrm{OR}}(\underset{\sim}{X}_N, \alpha_N), i = 1, \ldots, N) = 1 - \alpha \tag{1}$$

with $\alpha_N = 1 - (1 - \alpha)^{1/N}$ and $\alpha \in (0,1)$. This means that in a sample of size N really coming from the multivariate normal, with probability $1 - \alpha$ no observation will be identified as an α_N outlier.

We only consider affine equivariant estimates $\underline{m}$ and S, leading to identifiers with the same property.

3 The Maximum Asymptotic Bias

Davies and Gather (1993) consider, among others, especially two important types of worst-case performance criteria for univariate outlier identifiers: breakdown criteria (namely masking and swamping breakdown points) and the maximum asymptotic bias. The masking breakdown point of multivariate identifiers is investigated in detail by Becker and Gather (1997a). We will concentrate here on the maximum asymptotic bias, which expresses the behaviour of an outlier identification rule under a certain fixed amount of badly placed observations in a sample.

The definition of the maximum asymptotic bias of a multivariate outlier identifier is not straightforward. On the one hand, we have to determine the difference between the regions $\underline{\mathrm{OR}}$ and out, the complements of two ellipsoids or, equivalently, the difference between the ellipsoids $\mathbb{R}^p \backslash \underline{\mathrm{OR}}$ and $\mathbb{R}^p \backslash \mathrm{out}$ themselves. On the other hand, we want to use properties of the estimators $\underline{m}$ and S to derive properties of the resulting identifier $\underline{\mathrm{OR}}$. Therefore, we start with a definition of Tyler (1994, p. 1027), who introduces the *maximum bias* of a pair $(\underline{m}, S)$ of estimators of location and covariance *caused by* ε_m *corruption of a sample*, which in our notation with $\varepsilon_m = k/N$ reads

$$\begin{aligned} b(\frac{k}{N}, \underset{\sim}{x}_N; \underline{m}, S) \;\; = \;\; & \sup_{\underset{\sim}{x}_{N,k}} [\max\{\|S(\underset{\sim}{x}_N)^{-1/2}(\underline{m}(\underset{\sim}{x}_N) - \underline{m}(\underset{\sim}{x}_{N,k}))\|, \\ & \mathrm{tr}(S(\underset{\sim}{x}_N)S^{-1}(\underset{\sim}{x}_{N,k}) + S^{-1}(\underset{\sim}{x}_N)S(\underset{\sim}{x}_{N,k}))\}]. \end{aligned}$$

Here, $\underset{\sim}{x}_N$ denotes a sample of size N and $\underset{\sim}{x}_{N,k}$ is constructed from $\underset{\sim}{x}_N$ by replacing k observations out of N by arbitrary vectors.

Several steps are needed to adjust this definition to the necessities of outlier identifiers and to reach an analogous definition of the maximum asymptotic bias for the univariate case as given in Davies and Gather (1993).

First, instead of considering the values of $\underline{m}$ and S at the sample $\underset{\sim}{x}_N$ as a reference, we retain to the true parameters $\underline{\mu}$ and Σ. Second, we are only interested in the bias caused by "explosion" of the covariance part, thus we can reduce the term within the trace operator. Third, we do not regard samples with arbitrarily replaced observations but samples consisting of n regular observations from $N(\underline{\mu}, \Sigma)$ and an amount $k = N - n$ of δ_N outliers (for some $\delta_N \in (0,1)$). Therefore, the above supremum will be taken here over all combinations of the δ_N outliers. This is the same approach as in Davies and Gather (1993) who adapt a definition of Huber (1981, p. 12) to the outlier setting. Up to this point, our modified definition reads as follows:

$$b\left(\frac{k}{N}, \underset{\sim}{x}_N; \underline{m}, S\right) = \sup_{\underset{\sim}{x}_k^0} [\max\{\|\Sigma^{-1/2}(\underline{\mu} - \underline{m}(\underset{\sim}{x}_N))\|,\ \mathrm{tr}(\Sigma^{-1} S(\underset{\sim}{x}_N))\}], \tag{2}$$

where $\underset{\sim}{x}_N = (\underset{\sim}{x}_n^r, \underset{\sim}{x}_k^0)$ denotes a sample with n regular observations, $\underset{\sim}{x}_n^r = (\underline{x}_1^r, \ldots, \underline{x}_n^r)$, from $N(\underline{\mu}, \Sigma)$, and a number k of δ_N outliers, $\underset{\sim}{x}_k^0 = (\underline{x}_1^0, \ldots, \underline{x}_k^0)$. The supremum is taken over all $\underset{\sim}{x}_k^0 \in \mathrm{out}(\delta_N, \underline{\mu}, \Sigma)$, meaning that $\underline{x}_i^0 \in \mathrm{out}(\delta_N, \underline{\mu}, \Sigma) \forall i = 1, \ldots, k$.

This is still a definition of a property of a pair of estimators and not of an outlier identifier. So, the next steps have to adjust the definition to the situation of an identifier. The normalizing constants $c(p, N, \alpha_N)$ and $\chi^2_{p;1-\alpha_N}$ of the region $\underline{\mathrm{OR}}$ and of the region 'out' must be taken into account, because we actually compare $c(p, N, \alpha_N)S$ with $\chi^2_{p;1-\alpha_N}\Sigma$ instead of comparing S and Σ. For the second part of the above modification (2), we then get $\mathrm{tr}\left((c(p, N, \alpha_N)/\chi^2_{p;1-\alpha_N})\Sigma^{-1} S\right)$. From an inequality of Theobald (1975, p. 462) we see that

$$\mathrm{tr}\left(\frac{c(p, N, \alpha_N)}{\chi^2_{p;1-\alpha_N}} \Sigma^{-1} S\right) \geq \sum_{i=1}^{p} \frac{c(p, N, \alpha_N)}{\chi^2_{p;1-\alpha_N}} \frac{\xi_i}{\lambda_i},$$

where λ_i, ξ_i are the eigenvalues of Σ and S, respectively. Therefore, we can use the terms $(c(p, N, \alpha_N)\xi_i)/(\chi^2_{p;1-\alpha_N}\lambda_i)$ as characteristics of the bias of $c(p, N, \alpha_N)S$. On the other hand, $\sqrt{c(p, N, \alpha_N)\xi_i}$ and $\sqrt{\chi^2_{p;1-\alpha_N}\lambda_i}$ are equal to the lengths of the main axes of the ellipsoids $I\!\!R^p \backslash \underline{\mathrm{OR}}$ and $I\!\!R^p \backslash \mathrm{out}$ and thus can be interpreted in terms of the outlier identifier and the outlier region. To incorporate not only the lengths of the axes but also their orientations, we additionally introduce the eigenvectors of the matrices S and Σ, which results in defining the difference of the covariance part of the identifier and of the outlier region via the endpoints of the main axes by

$$\frac{1}{2} \sum_{i=1}^{p} \left(\left\|\underline{u}_i - \sqrt{\frac{c(p, N, \alpha_N)\xi_i}{\chi^2_{p;1-\alpha_N}\lambda_i}}\, \underline{v}_i\right\| + \left\|\underline{u}_i + \sqrt{\frac{c(p, N, \alpha_N)\xi_i}{\chi^2_{p;1-\alpha_N}\lambda_i}}\, \underline{v}_i\right\| \right) - p,$$

where $\underline{u}_i$ and $\underline{v}_i$ denote the eigenvectors of Σ and S to the eigenvalues λ_i and ξ_i, respectively.

The first part of (2) can be directly interpreted in terms of the identifier and the outlier region: $\|\Sigma^{-1/2}(\underline{\mu} - \underline{m}(\underset{\sim}{x}_N))\|$ describes the difference between the centers of the two ellipsoids $I\!\!R^p \backslash \underline{\mathrm{OR}}$ and $I\!\!R^p \backslash \mathrm{out}$, thus representing the location part of the difference.

The last step of our modification consists in summing up the differences instead of calculating their maximum. Otherwise two identifiers would have the same maximum asymptotic bias if they were based for example on identical estimators of covariance but different location estimates, and both location estimates were superior to the estimator of covariance. Differences between the two location estimators would not be taken into account. But in such a case the identifier using a better location estimator should have the smaller bias.

This leads to the following definition.

Definition 3.1 *The* maximum asymptotic bias *of an outlier identifier* $\underline{\mathrm{OR}}$ *is given as*

$$
\begin{aligned}
B(\underline{\mathrm{OR}}, \eta, \ ) := & \limsup_{N\to\infty} \Big(\sup_{\underset{\sim}{x}_k^0 \in \mathrm{out}(\delta_N, \underline{\mu}, \Sigma)} \big(\|\Sigma^{-1/2}(\underline{\mu} - \underline{m})\| \\
& + \frac{1}{2} \sum_{i=1}^{p} \Big\| \underline{u}_i - \sqrt{\frac{c(p, N, \alpha_N)\xi_i}{\chi^2_{p;1-\alpha_N}\lambda_i}}\, \underline{v}_i \Big\| + \Big\| \underline{u}_i + \sqrt{\frac{c(p, N, \alpha_N)\xi_i}{\chi^2_{p;1-\alpha_N}\lambda_i}}\, \underline{v}_i \Big\| - p \big) \Big),
\end{aligned}
$$

where $k := [\eta n]$, $0 < \eta < 1$, $N := n + k$, $\ := (\delta_i)_{i \in \mathbb{N}}$, $\delta_i \in (0,1)$. *Here,* $\|.\|$ *denotes the euclidean norm and* $[x]$ *is the integer part of* $x \in \mathbb{R}$. *The notation* $\underset{\sim}{x}_k^0 \in$ *out*$(\delta_N, \underline{\mu}, \Sigma)$ *is an abbreviation for "*$\underline{x}_i^0 \in$ *out*$(\delta_N, \underline{\mu}, \Sigma)$*, for all* $i = 1, \ldots, k$*".*

To find relationships between the bias of the estimators involved and the bias of the resulting outlier identifier, we also have to consider the maximum asymptotic bias for estimates of location and covariance. With the same notations as in the above definition, the *maximum asymptotic bias of a location estimator* $\underline{m}(\underset{\sim}{x}_N)$ for the parameter $\underline{\mu}$ is given by

$$
b(\underline{m}, \eta, \ ) := \limsup_{N\to\infty} \Big(\sup_{\underset{\sim}{x}_k^0 \in \mathrm{out}(\delta_N, \underline{\mu}, \Sigma)} \|\underline{m}(\underset{\sim}{x}_N) - \underline{\mu}\| \Big)
$$

and the *maximum asymptotic bias of an estimator* $S(\underset{\sim}{x}_N)$ *for the covariance matrix* Σ is defined as

$$
b(S, \eta, \ ) := \limsup_{N\to\infty} \Big(\sup_{\underset{\sim}{x}_k^0 \in \mathrm{out}(\delta_N, \underline{\mu}, \Sigma)} \|S(\underset{\sim}{x}_N) - \Sigma\|_2 \Big),
$$

where $\|.\|_2$ denotes the spectral norm of $\mathbb{R}^{p\times p}$. The choice of the spectral norm for measuring the distance between two matrices follows a proposal of Woodruff and Rocke (1993, p. 70).

4 Outlier Identifiers With Bounded Maximum Asymptotic Bias

With the above definitions, it is now possible to derive necessary conditions for outlier identifiers to have bounded maximum asymptotic bias.

Theorem 4.1 *For an outlier identifier $\underline{OR}$, based on estimators $\underline{m}$ and S, the following holds:*

(a) If $b(\underline{m}, \eta, \ ) = \infty$, then also $B(\underline{OR}, \eta, \ ) = \infty$.

(b) If $b(S, \eta, \ ) = \infty$, then also $B(\underline{OR}, \eta, \ ) = \infty$.

The proof is given in the Appendix. From Theorem 4.1 it is obvious that only estimators of location and covariance which both possess bounded maximum asymptotic bias yield outlier identifiers with the same property. Estimators with bounded maximum asymptotic bias are high-breakdown robust estimators as for example Rousseuw's (1985) MVE estimators or the MCD estimators proposed by Rousseeuw and Leroy (1987, p. 262) as well as the location-covariance S-estimates (Davies, 1987). In each case it can be seen from the proofs of the high breakdown points that those estimators have bounded maximum asymptotic bias if the proportion of outliers in the sample stays below the breakdown point.

For a sufficient condition, however, we need more.

Theorem 4.2 *Let $\underline{OR}$ be an outlier identifier as above with corresponding normalizing constant $c(p, N, \alpha_N)$. If the constant c fulfills the condition $c(p, N, \alpha_N) = \mathcal{O}(\chi^2_{p;1-\alpha_N})(N \to \infty)$, then from $b(\underline{m}, \eta, \ ) < \infty$ and $b(S, \eta, \ ) < \infty$ it follows that $B(\underline{OR}, \eta, \ ) < \infty$.*

We give the proof in the Appendix. At first sight, the condition on $c(p, N, \alpha_N)$ does not seem to depend on properties of the estimators $\underline{m}$ and S. But actually, under normalizing condition (1), the use of $\sqrt{N}$ consistent estimators guarantees that $c(p, N, \alpha_N) = \mathcal{O}(\chi^2_{p;1-\alpha_N})$. This is shown in detail in Becker and Gather (1997b); we just give a short sketch here. For $\sqrt{N}$ consistent estimators $\underline{m}$ and S, we have that, if $\underline{X}_1, \ldots, \underline{X}_N$ are i.i.d. according to

$N(\underline{\mu}, \Sigma)$, then $Y_i := (\underline{X}_i - \underline{m})^T S^{-1} (\underline{X}_i - \underline{m})$ are asymptotically χ^2_p distributed. Hence, we can derive the asymptotic distribution of $\max(Y_1, \ldots, Y_N)$, using results of Galambos (1987, p. 54, 102, 105) which show that the χ^2_p distribution lies in the (maximum) domain of attraction of the double exponential. From the normalizing condition (1) we see that $c(p, N, \alpha_N)$ is the $(1-\alpha)$ quantile of the distribution of $\max(Y_1, \ldots, Y_N)$. Therefore, for large N, we can approximate $c(p, N, \alpha_N)$ by the respective quantile of the double exponential, namely

$$c(p, N, \alpha_N) \simeq \chi^2_{p;1-1/N} - \frac{\ln(-\ln(1-\alpha))}{N f_{\chi^2_p}(\chi^2_{p;1-1/N})},$$

where $\alpha_N = 1 - (1-\alpha)^{1/N}$ and $f_{\chi^2_p}$ denotes the Lebesgue-density of the χ^2_p. Calculating $\lim_{N\to\infty} \; c(p, N, \alpha_N)/\chi^2_{p;1-\alpha_N}$ with the above relation gives a limiting value of 1. Thus, the following corollary holds.

Corollary 4.1 *If an identifier $\underline{OR}$ is based on $\sqrt{N}$ consistent estimators $\underline{m}$ and S for $\underline{\mu}$ and Σ, and if the normalizing condition (1) is used, then* $c(p, N, \alpha_N) = \mathcal{O}(\chi^2_{p;1-\alpha_N})(N \to \infty)$.

As shown above, the use of robust estimators of location and covariance with high breakdown points does not suffice to get an outlier identifier with bounded maximum asymptotic bias. Additionally, the used estimators should be $\sqrt{N}$ consistent. From the above mentioned examples both MCD and S-estimators fulfill this condition, but, in contrast to this, the MVE estimators do not, cf. Davies, 1992. We therefore rather recommend the use of MCD and S-estimators in multivariate outlier identification procedures.

Appendix: Proofs

Proof of Theorem 4.1

(a) Let $b(\underline{m}, \eta, \;) = \infty$. Then

$$\begin{aligned}
&\|\Sigma^{-1/2}(\underline{\mu} - \underline{m})\| \\
&+ \frac{1}{2} \sum_{i=1}^{p} \left(\left\| \underline{u}_i - \sqrt{\frac{c(p, N, \alpha_N)\xi_i}{\chi^2_{p;1-\alpha_N}\lambda_i}}\, \underline{v}_i \right\| + \left\| \underline{u}_i + \sqrt{\frac{c(p, N, \alpha_N)\xi_i}{\chi^2_{p;1-\alpha_N}\lambda_i}}\, \underline{v}_i \right\| \right) - p \\
&\qquad \geq \; \|\Sigma^{-1/2}(\underline{\mu} - \underline{m})\| - p \\
&\qquad \geq \; \frac{1}{\|\Sigma^{1/2}\|_2} \|\underline{\mu} - \underline{m}\| - p \quad = \quad \frac{1}{\sqrt{\lambda_1}} \|\underline{\mu} - \underline{m}\| - p, \qquad (3)
\end{aligned}$$

where λ_1 denotes the largest eigenvalue of Σ.

The above inequality remains valid for every $N \in I\!N$, if $\sup_{\underset{\sim}{x}_k^0 \in \text{out}(\delta_N, \underline{\mu}, \Sigma)}$ is taken on either side. Because of the condition on $b(\underline{m}, \eta, \ )$, it holds that

$$\limsup_{N\to\infty} \Big(\sup_{\underset{\sim}{x}_k^0 \in \text{out}(\delta_N, \underline{\mu}, \Sigma)} (\|\underline{\mu} - \underline{m}\|) \Big) = \infty.$$

Therefore, for every $R \in I\!R$ we can find $I_R \subseteq I\!N$, $|I_R| = \infty$, such that

$$\sup_{\underset{\sim}{x}_k^0 \in \text{out}(\delta_N, \underline{\mu}, \Sigma)} (\|\underline{\mu} - \underline{m}\|) > R \quad \forall\, N \in I_R.$$

From this, it follows immediately that

$$\limsup_{N\to\infty} \sup_{\underset{\sim}{x}_k^0 \in \text{out}(\delta_{N_0}, \underline{\mu}, \Sigma)} \Big(\frac{1}{\sqrt{\lambda_1}} \|\underline{\mu} - \underline{m}\| - p \Big) = \infty,$$

which together with (3) completes the proof.

(b) The proof is similar to part (a). Here, we use the inequality

$$\begin{aligned}
& \|\Sigma^{-1/2}(\underline{\mu} - \underline{m})\| \\
& + \frac{1}{2} \sum_{i=1}^{p} \Big(\Big\|\underline{u}_i - \sqrt{\frac{c(p, N, \alpha_N)\xi_i}{\chi^2_{p;1-\alpha_N}\lambda_i}}\, \underline{v}_i\Big\| + \Big\|\underline{u}_i + \sqrt{\frac{c(p, N, \alpha_N)\xi_i}{\chi^2_{p;1-\alpha_N}\lambda_i}}\, \underline{v}_i\Big\| \Big) - p \\
\geq \ & \sqrt{\frac{c(p, N, \alpha_N)}{\chi^2_{p;1-\alpha_N}}} \sum_{i=1}^{p} \sqrt{\frac{\xi_i}{\lambda_i}} - p \\
\geq \ & \sqrt{\frac{c(p, N, \alpha_N)}{\chi^2_{p;1-\alpha_N}\lambda_1}} \sqrt{\|S\|_2} - p.
\end{aligned} \tag{4}$$

With similar arguments as before, we find that

$$\limsup_{N\to\infty} \sup_{\underset{\sim}{x}_k^0 \in \text{out}(\delta_N, \underline{\mu}, \Sigma)} \|S\|_2 = \infty,$$

and this, together with (4) gives the stated result.

□

Proof of Theorem 4.2

We find

$$
\begin{aligned}
& \|\Sigma^{-1/2}(\underline{\mu} - \underline{m})\| \\
& + \frac{1}{2} \sum_{i=1}^{p} \left(\left\| \underline{u}_i - \sqrt{\frac{c(p, N, \alpha_N)\xi_i}{\chi^2_{p;1-\alpha_N}\lambda_i}}\, \underline{v}_i \right\| + \left\| \underline{u}_i + \sqrt{\frac{c(p, N, \alpha_N)\xi_i}{\chi^2_{p;1-\alpha_N}\lambda_i}}\, \underline{v}_i \right\| \right) - p \\
\leq \; & \frac{1}{\sqrt{\lambda_p}} \|\underline{\mu} - \underline{m}\| + \sqrt{\frac{c(p, N, \alpha_N)}{\chi^2_{p;1-\alpha_N}}} \sum_{i=1}^{p} \sqrt{\frac{\xi_i}{\lambda_i}} \\
\leq \; & \frac{1}{\sqrt{\lambda_p}} \|\underline{\mu} - \underline{m}\| + p \sqrt{\frac{c(p, N, \alpha_N)}{\chi^2_{p;1-\alpha_N}\lambda_p}} \max\{1, \xi_1\} \\
\leq \; & \frac{1}{\sqrt{\lambda_p}} \|\underline{\mu} - \underline{m}\| + p \sqrt{\frac{c(p, N, \alpha_N)}{\chi^2_{p;1-\alpha_N}\lambda_p}} \max\{1, \|S - \Sigma\|_2 + \lambda_1\}.
\end{aligned}
$$

Applying $\sup_{\underset{\sim}{x}_k^0 \in \mathrm{out}(\delta_N, \underline{\mu}, \Sigma)}$ on either side of the inequality and calculating the lim sup leads to

$$
\begin{aligned}
B(\underline{\mathrm{OR}}) \; \leq \; & \frac{1}{\sqrt{\lambda_p}} b(\underline{m}) \\
& + \frac{p}{\sqrt{\lambda_p}} \limsup_{N \to \infty} \left(\sqrt{\frac{c(p, N, \alpha_N)}{\chi^2_{p;1-\alpha_N}}} \sup_{\underset{\sim}{x}_k^0 \in \mathrm{out}(\delta_N, \underline{\mu}, \Sigma)} (\max\{1, \|S - \Sigma\|_2 + \lambda_1\}) \right) .
\end{aligned}
$$

Taking into account that $c(p, N, \alpha_N) = \mathcal{O}(\chi^2_{p;1-\alpha_N})$, which means that there exists some $M \in \mathbb{R}$ such that

$$
\frac{c(p, N, \alpha_N)}{\chi^2_{p;1-\alpha_N}} < M \quad (N \to \infty),
$$

we can further conclude that

$$
B(\underline{\mathrm{OR}}) \leq \frac{1}{\sqrt{\lambda_p}} b(\underline{m}) + \frac{p}{\sqrt{\lambda_p}} \max\{1, \sqrt{M}\} \max\{1, \lambda_1 + b(S)\} < \infty.
$$

$\square$

References

Barnett, V., Lewis, T. (1994), *Outliers in Statistical Data*, 3rd ed., Wiley, New York.

Becker, C., Gather, U. (1997a), The Masking Breakdown Point of Multivariate Outlier Identification Rules, *Technical Report 9/1997*, Sonderforschungsbereich 475, University of Dortmund.

Becker, C., Gather, U. (1997b), Convergence Rates in Multivariate Robust Outlier Identification, *Technical Report 10/1997*, Sonderforschungsbereich 475, University of Dortmund.

Davies, P.L. (1987), Asymptotic Behaviour of S–Estimates of Multivariate Location Parameters and Dispersion Matrices, *The Annals of Statistics*, **15**, 1269–1292.

Davies, P.L. (1992), The Asymptotics of Rousseeuw's Minimum Volume Ellipsoid Estimator, *The Annals of Statistics*, **20**, 1828–1843.

Davies, P.L., Gather, U. (1993), The Identification of Multiple Outliers, *Journal of the American Statistical Association*, **88**, 782–792.

Donoho, D.L., Huber, P.J. (1983), The Notion of Breakdown Point, in: Bickel, P.J., Doksum, K.A., Hodges, J.L. (eds.), *A Festschrift for Erich Lehmann*, Wadsworth, Belmont, CA, 157–184.

Galambos, J. (1987), *The Asymptotic Theory of Extreme Order Statistics*, 2$^{\text{nd}}$ ed., Krieger Publishing Company, Malabar, Florida.

Gather, U., Becker, C. (1997), Outlier Identification and Robust Methods, in Maddala, G.S., Rao, C.R. (eds.), *Handbook of Statistics, Vol. 15: Robust Inference*, Elsevier, Amsterdam, pp. 123–143.

Huber, P.J. (1981), *Robust Statistics*, Wiley, New York.

Rousseeuw, P.J. (1985), Multivariate Estimation with High Breakdown Point, in: Grossmann, W., Pflug, G., Vincze, I., Wertz, W. (eds.), *Mathematical Statistics and Applications*, **8**, Reidel, Dordrecht, 283–297.

Rousseeuw, P.J., Leroy, A.M. (1987), *Robust Regression and Outlier Detection*, Wiley, New York.

Theobald, C.M. (1975), An Inequality with Application to Multivariate Analysis, *Biometrika*, **62**, 461–466.

Tyler, D.E. (1994), Finite Sample Breakdown Points of Projection Based Multivariate Location and Scatter Statistics, *The Annals of Statistics*, **22**, 1024–1044.

Woodruff, D.L., Rocke, D.M. (1993), Heuristic Search Algorithms for the Minimum Volume Ellipsoid, *Journal of Computational and Graphical Statistics*, **2**, 69–95.